DOCUMENT RESUME

ED 453 247 TM 032 784

AUTHOR Brunsman, Bethany A.; Lukin, Leslie E.
TITLE Classroom Assessment: Possibilities for State Reporting of Student Proficiency.
PUB DATE 2001-04-00
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001).
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Comparative Analysis; Elementary Secondary Education; *School Districts; State Programs; *Test Construction; *Testing Programs
IDENTIFIERS Lincoln Public Schools NE; Nebraska

ABSTRACT

The state of Nebraska has adopted a unique approach to statewide reporting that requires school districts to develop district-level assessments of the state student content standards or district standards comparable in quality and to report the results to the state and public. This system allows districts to base their assessments on what they are teaching and to use creative approaches to assessment to measure outcomes that are not easily captured in paper-and-pencil tests. The state also hopes that districts will be able to use the same assessment results to enhance instruction. The response of the Lincoln Public Schools to the state requirements has included both paper-and-pencil tests and comparable classroom assessments. These comparable classroom assessments allow standardized judgments across classrooms through the use of scoring rubrics and teacher training. Teams of teachers trained in assessment development created district standards and assessments. Preliminary results suggest that judgments are fairly consistent among teachers. Additional studies are planned. Comparable classroom assessments seem to be viable and cost effective for measuring student achievement for the purposes of accountability and enhancement of instruction. (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

Classroom Assessment: Possibilities for State Reporting of Student Proficiency

ED 453 247

Bethany A. Brunsman and Leslie E. Lukin

Lincoln Public Schools

Paper Presented at the 2001 Annual Meeting of the

American Educational Research Association

Seattle, Washington

TM032784

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

B. A. Brunsman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

Abstract

Several people have argued recently that state testing and reporting policies that rely on statewide tests of content standards may not be working to improve student learning. The state of Nebraska has adopted a unique approach to statewide reporting that requires school districts to develop district-level assessments of the state student content standards or district standards comparable in quality and report the results to the state and public. This system allows districts to base their assessments on what they are teaching and to use creative approaches to assessment to measure outcomes that are not easily captured in paper-and-pencil tests. The state also hopes that districts will be able to use the same assessment results to enhance instruction. Lincoln Public Schools' response to the state requirements included both paper-and-pencil tests and comparable classroom assessments. These comparable classroom assessments allow standardized judgments across classroom through the use of scoring rubrics and teacher training. Teams of teachers trained in assessment development created district standards and assessments. Preliminary results suggest that judgments are fairly consistent among teachers. Additional studies are planned. Comparable classroom assessments seem to be viable and cost effective for measuring student achievement for both the purposes of accountability and enhancement of instruction.

Classroom Assessment: Possibilities for State Reporting of Student Proficiency

Introduction

According to a survey conducted by the Council of Chief State School Officers in 2000, 49 states had developed student content standards in some or all of the four core subject areas (English language arts, mathematics, science, and social science) taught in K-12 schools (CCSSO, 2000). All 49 states had adopted content standards in English language arts and math. These content standards specify what students at particular grade levels should know and be able to do in each content area. In an effort to improve student achievement by increasing accountability of schools, many states also measure students' progress toward meeting these standards at certain grade levels with statewide assessments and report the results publicly.

A number of people have argued recently, however that these state testing and reporting policies may not be working to improve student learning as well as policymakers intended. For example, Linn (2000) asked the question: "Have the assessment-based accountability models that are now being used or considered by states and districts been shown to improve education?" (p. 13). To answer this question, he compared trends over time for state assessments and for the National Assessment of Educational Progress (NAEP) in those states. The data suggest contradictory conclusions about changes in student achievement for the two sources of data. Gains on state assessments tend to be greater than gains on NAEP. Linn argues that the divergence of trends raises questions about the validity and generalizability of achievement gains on state tests.

Klein, Hamilton, McCaffrey, and Stecher (2000) took a closer look at math and reading results for students on the Texas Assessment of Academic Skills (TAAS), which is used to measure student progress on the Texas content standards as part of the state's accountability system. The researchers found that in Texas, the results of the TAAS and the NAEP are somewhat inconsistent. Between 1994 and 1998, TAAS scores increased dramatically and the gap among racial and ethnic groups diminished. Gains on the NAEP in Texas during that time period, however, were much more modest and the gap among groups based on race and ethnicity increased slightly. In addition to the misleading gains in achievement, Klein et al. discuss a number of negative unintended consequences of the accountability program including narrowed curriculum, inappropriate preparation, increases in student retention at certain grade levels, increases in student dropouts, and increased exclusion of special needs students from testing.

Why does increased accountability not lead to better education? Stiggins (1999) asked the same question in another way: "When unsupported and angry teachers rely on potentially counterproductive strategies to teach students, who regard academic success as beyond their reach and who have stopped caring, is the result likely to be significant school improvement? Is the result likely to be an increase in the proportion of our students who meet state or local academic standards?" Stiggins argues that focusing assessment resources on accountability systems increases the anxiety of both teachers and students without giving them the tools to improve learning. This situation leads to frustration on the part of both teachers and students. Stiggins proposes that a more balanced approach to assessment would be a better approach to improving education. State- or district-level assessments for accountability combined with resources to improve assessments that provide information to students and teachers about instruction and learning would provide the means to achieve improved education.

Similarly, Popham (1999) contends that large-scale assessment programs are too much focused on accountability purposes and provide very little useful information to facilitate instruction and student learning. He recommends the following changes in large-scale assessments to allow them to continue to serve accountability purposes while also contributing to instruction:

1. Test development efforts include people with experience teaching in classrooms,
2. Tests be designed to measure knowledge and skills that are important and teachable,
3. Assessment domains be clearly and specifically defined before samples of items are chosen to measure them, and
4. States and districts stop using national standardized tests to evaluate education quality.

Linn (2000) adds another assertion; he argues as other have that what is measured in an assessment needs to be carefully determined because content areas and subareas that are assessed are emphasized in instruction, whereas other content may not be taught at all. State content standards for students often contain outcomes that are not easily measured (or measured at all) by standardized paper-and-pencil tests. English language arts standards, in particular, often

contain writing, listening, and speaking standards. Science standards may require students to demonstrate the ability to use equipment safely or the process of inquiry. For the types of products contained in many state standards in these content areas, performance assessment is the most appropriate method. In listening and speaking we can measure prerequisite knowledge and skills using multiple-choice and essay items, but we cannot directly measure a student's ability, for example, to deliver an oral presentation using these methods. Writing standards often contain language requiring students to improve writing over time through revision. This aspect of writing is difficult to measure in one or two sittings.

Many state assessments include short answer or extended response items and several states have added projects or portfolios of student work to statewide testing efforts to measure product outcomes (CCSSO, 2000). Centralized scoring of open-ended items and performance assessments is very expensive, however. As a result, performance assessments at the state level must contain very few items to be affordable. States may draw conclusions about student proficiency in writing, for example, on the basis of student responses to as few as one writing prompt (e.g., the Missouri high school state writing assessment).

Research in several content areas suggests that student performance within a domain may vary significantly from one task to another (Dunbar, Koretz, & Hoover, 1991). The more heterogeneous the content, the more variability that was found. Dunbar, et al.'s review of studies in writing suggests that student scores on essays written within one mode of discourse (e.g., narrative, persuasive) were only moderately correlated. The correlations were even smaller across modes of discourse. Estimates of the number of tasks required to reliably assess a student's proficiency in one content area range from 8 to 20 (Herman, 1997).

One way to reduce the number of required tasks is to more narrowly specify the content domain. Measuring "persuasive writing" requires fewer tasks than does measuring "writing." If we want to measure writing proficiency, however, measuring persuasive writing will not allow us to make valid inferences about student proficiency. Moreover, if we only measure persuasive writing in a moderate or high stakes statewide assessment, schools may only teach persuasive writing.

How can we measure student achievement in content areas like speaking and writing without compromising either the complexity of the content domain or the validity of inferences? One solution may be to use comparable classroom-based assessments. Lincoln, Nebraska Public

Schools decided to pursue comparable classroom assessments as a cost-effective way to meet state reporting requirements.

The state of Nebraska has a unique approach to statewide standards-based assessment compared with other states. Rather than adopting one or several statewide tests to measure student proficiency in each set of content area standards, Nebraska requires school districts to develop or adopt district-level assessments of either the state standards or district standards comparable in quality. The main purpose of this plan is to improve student achievement by allocating more resources to classroom assessment, while still collecting information that can be used in state-level and local policy decisions. The rationale is that locally-developed assessments will be more aligned to district curricula and will provide results that can be used by teachers to enhance instruction.

The Nebraska approach will involve compromises. The purposes of holding schools accountable and providing useful information for instructional decisions are not easily fulfilled by the same assessment. Statewide standardized assessments generally provide good policy-level data. These types of assessments are usually too infrequent and too broad in content to be used in day-to-day instructional decisions. Classroom assessments, on the other hand, provide information to teachers, parents, and students about day-to-day learning and can be used to adjust instruction, but they often lack the comparability of scores of large-scale standardized tests. As a result, they often provide little useful data for policy makers.

Popham (1999) contends that these two assessment purposes, accountability and adjustment of instruction, are not "inherently contradictory" (p. 15). He goes on to say

> It simply suggests that in order for large scale assessors to accomplish more in the instructional realm, without diminishing the accountability virtues of their assessments, substantially more energy must be devoted to the *instructional* side of the enterprise. We are not dealing with a zero-sum game in which increased attention to instruction requires decreased attention to accountability. Given sufficient asse. sment cleverness, this is a situation permitting simultaneous cake-having and cake-eating. (p.15)

The challenge is to combine standardized and classroom-based assessments without losing either consistency across classrooms or validity of results for instruction. Classroom assessments can be comparable across classrooms if teachers are trained in assessment and provided with materials that allow them to make standardized judgments about student performance. The key to comparability is reliability of teacher judgments. Another very important factor is the validity of those judgments. If the results of one assessment are going to be used for two different purpose (instructional decisions and accountability to the public), those results must be valid for both purposes. The assessments must be of high quality and very closely aligned with both the district curriculum used in classrooms and the content standards student achievement will be compared against. Data collection must be ongoing for use by teachers and reportable at the end of a given period. Students must have multiple opportunities of various types to demonstrate proficiency. In writing, teachers need to collect and evaluate multiple drafts of student work to measure revision.

By giving school districts flexibility about how student proficiency will be measured, the state of Nebraska has provided a unique opportunity to districts to measure and report student proficiency on all of the state (or district) adopted content standards. Because districts are not limited to multiple-choice, or even paper-and-pencil tests, student achievement in listening, speaking, and writing can be more fully measured at a local level than would be possible on a statewide test. Moreover, because the assessments do not need to occur in one sitting at a particular time, teachers and students can look at changes in proficiency over time and use this information to improve student learning.

Comparable Classroom Assessments: The Lincoln Public Schools Solution

Teachers, curriculum specialists, and assessment specialists at Lincoln Public Schools (Lincoln, Nebraska) worked together to develop a district assessment system intended to measure all of the district standards (which have been approved by the state as equally rigorous as the state standards). The assessment system includes both paper-and-pencil, standardized tests and comparable classroom-based assessments. The scores from the classroom assessments are based on standardized teacher judgments. The following description of the locally-developed assessments for English language arts focuses on the process of development and validation of the assessments.

Development

The district adopted a process of developing assessments in which teams of teachers would write the district standards and the assessments with support from the assessment and curriculum specialists. A lead teacher and teams of two to four teachers at each of the grade levels (4, 8, and 11) were selected to develop the district standards and assessments. According to this plan, experts in both the ELA content and in classroom realities would be involved in assessment development. The teachers worked part-time in the district office under their teacher contracts for one to two years. Some of the work was also completed in the summer.

All of the teachers involved in assessment development participated in 20 hours of assessment training. The training included the following topics:

- overview of assessment model
- attributes of quality assessment
- assessment development process
- assessment methods
- writing items/tasks
- reviewing items/tasks
- assessment bias
- assessments currently in use in the district
- overview of state English language arts standards

Many of the teachers involved in assessment development also participated in assessment literacy learning teams with other teachers in the district based on a model suggested by the Assessment Training Institute (Assessment Training Institute, 2000).

Following the assessment training, the first step in developing standards in ELA was to study the alignment of district curriculum objectives at grades 4, 8, and 11 with the state ELA standards. The teams of teachers reviewed the state standards and reworded them to align with district objectives. They then checked the match between the newly-written district standards and the norm-referenced standardized test currently used in the district and textbooks and other curriculum materials. They produced a number of documents to explicate these connections and allow teachers in the district to fully understand the content of each standard. The district

standards in ELA for grades 4, 8, and 11 were approved by the Nebraska Board of Education as equally rigorous to the state standards.

After developing the district standards, the teams of teachers decided what types of achievement targets needed to be measured and what method was most appropriate for each outcome. After considering both the match between the content and assessment method and available district resources, two types of assessments seemed most appropriate: standardized, selected-response tests and comparable classroom assessments. The comparable classroom assessments were performance assessments that would occur in classrooms. Teacher judgments would be standardized through the use of scoring rubrics, teacher training, and suggested activities for measuring the standards.

In grade four, the teachers chose to measure reading comprehension and vocabulary with multiple-choice tests. This type of test is efficient and fits with the knowledge and skills covered in most of the grade four reading standards. For the standards in listening and speaking and the standard related to personal reading, the fourth-grade teachers decided to use comparable classroom assessments[1]. These outcomes were not easily measured by selected-response formats. Standardized (one-time) performance assessments conducted at a similar time throughout the district were rejected for several reasons. First, that kind of assessment would require collection of videotapes or written documentation of performances and large-scale scoring by trained teachers at the district level. The costs both monetary and in other resources (e.g., people) did not seem justified. Second, standardized performance assessments would be limited in the number of samples of student performance they could include, again, because of costs and teacher and student time. Third, centrally-scored performance assessments would be of less use to teachers and students than classroom-based assessments because of the time it would take to return scores to teachers and students.

In grades 8 and 11, the teams of teachers decided to measure all of the standards with comparable classroom assessments for the same reasons discussed at grade four. Selected-response tests were discarded as an alternative for measuring reading because the reading

[1] Assessments for the standards in writing at grade four have not yet been developed. The state of Nebraska does not require districts to report student progress in writing beyond the results of the statewide writing assessment (the only statewide test in Nebraska). Comparable classroom assessments in fourth grade writing will be developed at a later date for use in instruction, school improvement, and other purposes.

standards at grades 8 and 11 focus on analysis of literature, an outcome that does not lend itself well to selected-response formats.

After deciding on the appropriate assessment methods, the teams of teachers at each grade began developing items, scoring rubrics, sample activities, and other supportive materials. The multiple-choice tests for fourth grade are based on tables of specifications containing the standards and matched district objectives. Twice as many items as were needed to measure each standard were developed. The reading passages on the reading comprehension test were selected based on both quantitative and qualitative readability analyses.

The comparable classroom assessments consist of scoring rubrics containing four levels of student achievement. Teachers report a holistic score for each standard. The rubrics for grades 8 and 11 contain four levels that mirror the levels in the state reporting system: beginning, developing, proficient, and advanced (see Appendix A for an example). The grade four rubrics contain three levels: developing, proficient, and advanced (see Appendix B for an example). Suggested activities to measure student performance accompany the rubrics. All activities are cross-referenced by the standards they may be used to measure. Some activities may be used with more than one rubric and more than one standard. The assessments allow teachers to choose activities that best fit with their teaching styles, with student learning styles, and with the particular materials (e.g., books) they use in their classrooms. Teachers are encouraged to share scoring rubrics with students.

Items, rubrics, activities, and supporting materials were then reviewed by groups of two to ten teachers in the grade levels for which they were designed. These teachers were briefly trained to review items and other materials in terms of content, form, match with curriculum, and appropriateness for students.

The teams of teachers revised the materials based on suggestions made by reviewers. Following the review, all materials were pilot tested in a sample of fourth, eighth grade and high school classrooms. Teachers used the activities, rubrics, and materials with students. They collected student data and made comments on the materials.

The assessment specialists computed item statistics for each of the multiple-choice items. The materials were revised a second time based on the results of the pilot test and items for the multiple-choice tests were selected and assembled into final forms of the tests.

The district has standing elementary and secondary bias review committees for testing materials. Teachers and administrators on these committees have been trained to detect bias in test items and materials. Ten to fifteen committee members reviewed the materials at each grade level for both offensiveness and unfair penalization of students based on socioeconomic status, race, ethnicity, culture, religion, or gender. The teams of teachers revised materials if more than 20% of reviewers flagged an item, rubric, or activity as biased.

Training

The teams of teachers worked with curriculum and assessment specialists to develop training materials for use in district-wide staff development. Staff development on the standards and assessments was critical for two reasons. First, it was important that teachers be familiar with the district standards so that they follow the district curriculum and students have opportunities to learn the knowledge and skills covered in the standards. Second, a major component of the comparable classroom assessments is standardized teacher judgments. Training is one way to increase consistency in judgments across teachers and students.

Table 1 contains information about the training provided to teachers related to the ELA standards and comparable classroom assessments. All teachers at grades 4 and all English Language Arts and Oral Communications teachers at grades 7-8 and 9-12 in the district participated in the training. Additionally, many resource teachers and special education co-teachers attended two hours of voluntary training related to the ELA standards and assessments.

Student Data Collection

All students at grades 4, 8, and 10 and selected ninth and eleventh graders are currently participating in the assessments. Fourth-grade students took the selected-response tests in reading comprehension and vocabulary in March of 2001. Teachers will use scannable forms to report each eighth-grade student's scores on the ELA standards based on the rubrics for the comparable classroom assessments in early May 2001. Scores for high school students are collected in English classes for reading and writing and in Oral Communication classes for listening and speaking. Because high school students may enroll in different classes or the same class with different teachers in first and second semester, teachers report scores for students at the end of each semester. The decision to collect scores for ninth and tenth grade students was based on the fact that students may take Oral Communication at any time in grades 9 through 12 and may not be enrolled in an English class in eleventh grade. The assessment specialists will

keep a database to track student progress toward the standards. In most cases, a student's most recent scores will be used for local decision making and to report to the state. A decision will be made in the district about what to do about scores for students who are reported as "proficient" by one teacher and later reported as "not proficient" by another teachers based on the numbers of these instances and the circumstances surrounding them.

Standard Setting Studies

A standard setting study will be conducted for the selected-responses tests in reading comprehension and vocabulary after the tests have been scored. For each standard, three cut scores will be determined, which will provide four categories of student performance: beginning, emerging, proficient, and advanced. According to the state rules, student performance with either be reported based on these four categories or as "proficient" and "not proficient".[2] The standard setting study will include two different methods: Modified Angoff and Borderline Group.

The rubrics for the comparable classroom assessments were designed to align with the four state reporting categories. As a result, standard setting studies in the classical sense would not be appropriate. What is needed to set the "standard" is to get all of the teachers at a particular grade level to agree on a level of performance, as described in the rubrics, that defines each category. This agreement or standardization of teacher judgments will be achieved through training and the use of student exemplars. Some of this training began this year with respect to the student presentation standard as described in the inter-rater reliability studies in the next section. Training will continue next year when more student exemplars will be available for distribution.

Validity Studies

In addition to the content-related evidence of validity of the scores that has already been presented, the following studies have been conducted or are planned to document evidence of reliability and validity of scoring and scores for the classroom assessments.

- A review of the assessment materials by teachers and college faculty outside of the school district to verify the match between the assessments and the standards,

[2] For schools receiving Title I funding, the state requires that achievement data be reported in the four specified categories. Data for other schools may be reported as numbers of students who are "proficient" and "not proficient."

coverage of the standards by the assessments, appropriateness of the assessments for students;

- A survey of teachers at appropriate grade levels to determine where the assessment content is addressed in lesson plans (opportunity to learn);
- Teacher training in scoring the comparable classroom assessments with the use of student exemplars;
- Studies of inter-rater reliability on the comparable classroom assessments;
- Analysis of internal consistency reliability for selected-response tests; and
- Comparisons of scores for reading, writing, listening, and speaking with scores on the Metropolitan Achievement Test (MAT), the district Graduation Demonstration Exams, the tenth grade PLAN, course grades, and the statewide writing assessment (administered in February at grades 4, 8, and 11).

Because the scores from all of the assessments are not yet available, only preliminary studies have been completed. In January, elementary and high school teachers participated in a scoring training for the student presentation rubrics. Seven videotaped student presentations were selected at grades 4 and 11. All of the fourth grade teachers in the district and a sample of high school Oral Communications teachers participated in the training. Participants watched the videotaped presentations and made preliminary judgments based on the scoring rubrics, which they had previously been given to use with their own students. After the judgments and notes on individual presentations were collected, participants discussed the scoring rubrics and the exemplars in detail and carefully defined characteristics of students at each of the scoring points.

Tables 3 and 4 contain the summarized judgments of teachers collected during the scoring training. Because these judgments were collected before the discussions of the scoring rubrics and student scores, they reflect the judgments teachers made on the scoring rubrics with very little training.

Conclusions

Comparable classroom-based assessments seem to be a viable and cost effective alternative for measuring student achievement when the purposes for assessment are both accountability and enhancement of instruction. The preliminary results from the Lincoln Public Schools assessment model suggest that they will provide reliable and valid data both for statewide accountability purposes and for classroom instructional decisions. Because they were

developed specifically to measure the district standards (which are aligned with state standards), they are a better measure of student proficiency for these purposes than many nationally-available standardized tests. The fact that they are based on multiple samples of student performance strengthens inferences to these heterogeneous content domains. In addition, their ongoing nature makes them much more instructionally relevant than most statewide standardized assessments. Although comparable classroom assessments may be slightly lower in reliability than are standardized paper-and-pencil tests, the direct relationship of the comparable classroom assessments to the district curriculum will make inferences based on scores more valid for both accountability and instructional purposes.

The Lincoln Public Schools model has some clear advantages. As both Stiggins (1999) and Popham (1999) suggest, it moves the focus from district assessments solely for accountability purposes to a balanced, shared focus of accountability and instruction.

The process of developing assessments was consistent with Popham's recommendations for developing a district assessment that will contribute to instruction. District teachers of English language arts wrote the district standards and developed the assessments. They developed district standards directly based on the district curriculum and selected assessment methods that were aligned with these achievement targets. Finally, they developed a number of documents that clearly specify the content and skills covered in the standards and the assessments.

A district assessment system with a balanced focus on both accountability and instruction is important for improving student achievement. As Stiggins (1999) argues, an assessment system designed only for accountability does not provide any data or tools for teachers and students to actually improve achievement. By moving to a shared assessment focus, Lincoln Public Schools and the state of Nebraska is providing teachers and students with the resources they need to increase student learning.

References

Assessment Training Institute. (2000). Learning Team Guide for Student-Involved Classroom Assessment, 3rd Ed. Portland, OR: Author.

Council of Chief State School Officers. (2000). Key state education policies on K-12 education: 2000: Time & Attendance, Graduation, Content Standards, Teacher & School Leader Licensure, Student Assessment. Washington, DC: Author.

Dunbar, S. B., Koretz, D. M., & Hoover, H. D. (1991).Quality control in the development and use of performance assessments. Applied Measurement in Education, 4, 263-371.

Herman, J. L. (1997). Large-scale assessment in support of school reform: Lessons in the search for alternatives (CSE Tech. Rep. No. 446). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Klein, S. P., Hamilton, L. S., McCaffrey, D. F., & Stecher, B. M. (2000). What do test scores in Texas tell us? Education Policy Analysis Archives, 8 (4).

Linn, R. L. (2000). Assessments and accountability, Educational Researcher, 29 (2), 4-16.

Popham, W. J. (1999). Where large scale educational assessment is heading and why it shouldn't. Educational Measurement: Issues and Practice, 18 (3), 13-17.

Stiggins, R. J. (1999). Assessment, student confidence, and school success. Phi Delta Kappan, November 1999, 191-198

Stiggins, R. J. (1997). Student-centered classroom assessment (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Author Note

We would like to thank the following teachers for their invaluable work in developing district English language arts standards and assessments: Terry Abrahams, Elaine Caster, Teresa Eckhout, Pam Gannon, Vicky Harris, Michelle Hohenfeldt, Pat Kaltenberger, Terry Keefe, Jana Maddox, David Smith, and Bev Westerberg.

Table 1.

District-wide Staff Development Related to the ELA Standards and Assessments

Topic	Teacher Grade Level(s)		
	4	7-8	9-12
District ELA Standards	2.0 hours	3.5 hours	1.5 hours
Assessments: Review and Practice	1.5 hours	3.5 hours	7.0 hours
Scoring: Student Presentations	3.5 hours		3.5 hours[a]

[a]A sample of high school Oral Communication teachers participated in this part of the training.

Table 2

Pre-training Inter-rater Reliability for Student Presentation Scores: Grade 4

Student	Score			Percent Exact Agreement
	Emerging	Proficient	Advanced	
1 (n=108)		103	5	95%
2 (n=109)	92	17		84%
3 (n=107)	1	66	40	62%
4 (n=107)	71	36		66%
5 (n=35)		8	27	77%
6 (n=36)		30	6	83%
7 (n=27)	2	19	6	70%

Table 3

Pre-training Inter-rater Reliability for Student Presentation Scores: High School

Student	Score				Percent Exact Agreement
	Beginning	Emerging	Proficient	Advanced	
1 (n=12)	3	9			75%
2 (n=12)	3	8	1		67%
3 (n=12)			12		100%
4 (n=11)		2	9		82%
5 (n=12)	9	3			75%
6 (n=12)		3	9		75%
7 (n=11)			7	4	64%

Appendix A:

Lincoln Public Schools High School Oral Presentation Rubric

12.3 SPEAKING--Oral Presentation Rubric

12.3.2 By the end of the twelfth grade, students will make oral presentations/public addresses that demonstrate appropriate consideration of audience, purpose, and information to be conveyed.

Scale: 0=Not Participating* 1=Minimal Evidence 2=Gaining Proficiency 3=Proficient 4=Exemplary 5=Not Assessed**

1 Minimal Evidence	2 Gaining Proficiency	3 Proficient	4 Exemplary
♦ no purpose established; no attempt to gain audience attention	♦ purpose is difficult to discern; little attempt to gain audience attention	♦ purpose is communicated; uses a technique to gain audience attention	♦ purpose is clear and smoothly incorporated into introduction; engages audience
♦ content is incomplete, disconnected and disorganized; lacks research or support	♦ content present but not developed, inadequate research or support	♦ relevant information that is developed and organized; adequate research and support	♦ relevant information that is fully developed, clearly organized with strong transitions and word choices; thoroughly researched and supported with multiple examples
♦ no clear sense of ending	♦ vague or trite sense of ending	♦ conclusion connects to introduction and body; creates a sense of ending	♦ conclusion seamlessly connects with introduction and body; creates a clear sense of ending
♦ is disconnected from audience	♦ aware of audience reaction but does not adjust	♦ aware of audience reaction and attempts to adjust to their needs	♦ highly aware of audience and can easily adjust to their needs and feedback
♦ may mumble or deliver speech in monotone; lacks energy; little eye contact; unaware of body language; language may be inappropriate	♦ may deliver speech in a clear voice but with no inflection; may lack energy; attempts eye contact; some awareness of body language; language may be inappropriate	♦ delivery is clear, varied, and energetic; eye contact is adequate; aware of body language; uses appropriate language	♦ delivery is articulate and energetic; inflection is used to underscore the message; eye contact is strong; uses body language for emphasis; uses appropriate language

***Student may have been absent or decided not to participate in assessed activity.**
****Skill not addressed in the assessed activity.**

Appendix B:

Lincoln Public Schools Grade 4 Oral Presentation Rubric

Fourth Grade DLO
4.8.2 PRESENTATIONS Speaks before a group to express or defend an opinion, present information, give directions, or share a book, story, or poem

PRESENTATIONS	Standard Not Yet Met	Standard Met	Standard Exceeded
4.8.2a Organization	✏presentation is unorganized lacking parts such as beginning, middle, or end	✏presentation is organized with a beginning, middle, and end	✏presentation is well organized with a beginning, middle, and end; transitions are smooth and natural
	✏purpose or theme is not apparent	✏there is a clear purpose or theme with main ideas and some details	✏there is a strong purpose or theme, with clear main ideas and vivid supporting details
	✏information is incomplete and may be inaccurate	✏information is accurate with some detail	✏information is complete, accurate, and includes detail
4.8.2c Eye Contact	✏makes very little or no eye contact with audience	✏attempts to make eye contact with audience	✏makes eye contact with audience naturally and often
4.8.2c Pace	✏rate of speech is too fast or too slow, distracting audience	✏rate of speech is not too fast or too slow, not distracting audience	✏rate of speech is slow enough for audience to think and respond and fast enough to hold their attention
4.8.2c Volume	✏words are inaudible	✏words are heard clearly	✏all words are heard clearly with varied tone of voice
4.8.2b, 4.8.2c Enunciation, Fluency and Expression	✏many words mispronounced	✏words pronounced correctly	✏all words pronounced correctly
	✏lacks expression, does not hold the attention of the audience	✏expression holds the attention of the audience, but may not yet be natural	✏ expression is natural and makes presentation exciting, holding the attention of the audience
4.8.2c Body Language	✏appears unsure, may wiggle or fidget	✏appears poised and prepared, without distracting gestures	✏appears confident and prepared using appropriate body language